Application: Enabling Biomedical Science with Common Workflow Language

Sarah Wait Zaranek - swz-grants@curii.com
Essential Open Source Software for Science- Cycle 2

Summary

ID: EOSS2-0000000241
Last submitted: Feb 4 2020 07:15 PM (EST)

Applicant Details Part 1

Completed - Apr 30 2020

Applicant Details Part 1

Please complete the following information for the Applicant (required):

The Applicant is the individual submitting the application who will act as the main person responsible for the application and point of contact for the application. Information about other Key Personnel on the proposal should be entered where requested in the Software Project Details part of the application.

All fields are required.

1. First Name	Sarah Wait
2. Last Name	Zaranek
3. Email	swz-grants@curii.com
4. Current Employer	Curii Corporation
5. Title/Position	CEO and Chief Scientist
6. Country of Residence	United States
7. Applicant Organization	Curii Corporation

Applicant Details Part 2

Completed - Feb 2 2020

Applicant Details Part 2

Please complete the following information for the Applicant. The Applicant is the individual submitting the application who will act as the main person responsible for the application and as its point of contact for the application.

Please note demographic information will not be used as a basis for review.

1. Highest degree:

1a. Degree (required)	PhD or equivalent
1b. Year (optional)	2005

2. Additional degrees (optional):

	Degree	Year
2a. Additional degree and year	MSc Geological Sciences	2002
2b. Additional degree and year	MSc Applied Mathematics	2002
2c. Additional degree and year	BSc Physics and Applied Mathematics	1999

3. Gender (optional):

Woman / Female

4. Short narrative biography of the applicant (required):

Maximum of 100 words

Dr. Sarah Wait Zaranek is a versatile computational scientist and leader with more than twenty years of industry and academic experience in parallel computing and distributed systems. She is the chief scientist at Curii Corporation which she founded with Dr. George Church of Harvard. At Curii she collaborates with highly interdisciplinary teams, from many different organizations, that use Common Workflow Language to process and gain insights from genomic and other biomedical data. Her central focus is fostering a vibrant ecosystem of data scientists and others who share common open-source tools and public data to advance precision medicine.

Organization Details

Completed - Feb 4 2020

Organization Details

Please complete all requested information for the organization that would be directly receiving and distributing funding if a grant is successfully awarded (e.g. academic institution, fiscal sponsor). Note that this may be different from the Applicant's main affiliation.

1. Type of Organization (required):

Industry/company

2. Organization (required):

2a. Organization Name	Curii Corporation
2b. Address	212 Elm Street
2c. City	Somerville
2d. State/Province	MA
2e. Country	United States
2f. Tax ID (9-digits; format: XX-XXXXXXX; enter 44-4444444 if not applicable)	820858075

3. Organizational/Administrative Contact (required):

List the name and contact information for the administrative contact to discuss additional information needed, if selected for award.

3a. First Name	Jeremy
3b. Last Name	Walter
3c. Title/Position	Research Administrator
3d. Email	jeremy-grants@curii.com

4. Signing Official (required):

List the name and contact information for the person authorized to sign on behalf of your organization.

4a. First Name	Alexander
4b. Last Name	Zaranek
4c. Title/Position	Vice President
4d. Email	awz-grants@curii.com

5. Press Contact / Public Relations Official (required):

List the name and contact information for the person to discuss press releases and media.

5a. First Name	Marc
5b. Last Name	Rubenfield
5c. Title/Position	Director Business Development
5d. Email	mrubenfield@curii.com

6. Institutional Approval Form (required):

This form should be reviewed and signed by a person authorized to sign on behalf of your organization or fiscal sponsor (whichever organization that will be receiving the grant funds, if the application is selected for funding), agreeing to the stated institutional and applicant requirements and commitments on data, resource sharing, and publication policies, as well as endorsing/verifying your application materials and confirming their ability to receive funding for the proposal and disperse funds to other projects named on the application. Applicants may be asked to provide additional information on the governance of the project and its ability to receive and use the funds, if successfully awarded. If you do not have an organization identified to receive funds, the applicant should sign the form and note on the form that a fiscal sponsor has not been identified.

Upload as a single PDF. ***This upload is not designed to support encrypted documents or digital signatures; please sign, scan and upload this form as a PDF.***

Inst_Approval_Form_Sept_2019 (1).pdf

Filename: Inst_Approval_Form_Sept_2019 (1).pdf **Size:** 339.9 kB

Proposal Details

Completed - Feb 4 2020

Proposal Details

Please complete the following proposal information. ***All sections are required.***

1. Proposal Title: **Enabling Biomedical Science with Common Workflow Language**

To edit your proposal title, navigate to the main page; click on the three dots to the right of the application title (next to the Preview link); and select Rename from the dropdown menu. Proposal title is limited to a maximum of 75 characters, including spaces.

2. Did you previously apply for funding for this or a related proposal under the EOSS program?

Yes- please specify application number, e.g. EOSS-0000005345: EOSS-0000000054

3. Proposal Purpose (required):

Limit to one sentence (maximum of 255 characters, including spaces)

Enable portability of complex biomedical workflows across different clouds and on-premise environments via better documentation, community support, and tooling for Common Workflow Language (CWL) and Arvados with examples from the Personal Genome Project.

4. Amount Requested (required):

Enter total budget amount requested in USD, including indirect costs; this number should be between $50k and $250k total costs and match total described in the Budget Description. Enter whole numbers only (no dollar signs, commas, or cents)

249976

5. Abstract/Proposal Summary (required):

Provide a short summary of the application (maximum of 250 words)

Common Workflow Language (CWL) is an essential technology for describing biomedical analysis workflows used by many organizations. Embracing CWL as a workflow standard enables reproducible research. This ensures that scientific work can be understood, independently verified, and built upon in future work. Improvements to CWL benefit multiple organizations and workflow systems that use the CWL standard. For these organizations, there is an ongoing need to train new users and develop new workflows with CWL. CWL faces barriers to wider use and adoption due to a perceived learning curve and usability challenges. Our proposal outlines a plan to help reduce those barriers through community engagement and improved usability. It covers documentation updates, community feedback aggregation and response, specification development, and tool creation. We propose funding Peter Amstutz to lead this effort alongside the PI Sarah Zaranek and a "community engineer." Peter is a leader in the development of CWL, "cwltool", "cwl-ex" and Arvados. He has demonstrated the ability to serve the needs of the community at large. He is familiar with "wearing many hats", including developer, technical writer, community relations, and project evangelist. Our overarching goals is to make it easier for users to create CWL workflows and run them at scale using any conformant CWL platform (drawing heavily on examples from the open-source Arvados platform developed by Curii.) Throughout the effort we we will use real world public domain (CC0) data and workflows from the global Personal Genome Projects where we have contributed since 2005.

6. Work Plan (required):

Provide a description of the proposed work the applicants are requesting funding for, including resources the applicants will provide that are not part of the requested funding. For software development related work (e.g., engineering, product design, user research), specify how the work fits into the existing software project roadmap. For community outreach related activities (e.g., sprints, training), specify how these activities will be organized, the target audience, and expected outcomes (maximum of 750 words)

CWL is a unique multi vendor standard for describing biomedical analysis workflows used by many organizations including Curii Corporation, Seven Bridges Genomics, Cincinnati Children’s Hospital, the Broad Institute, Argonne National Labs, Roche Pharmaceuticals, Duke University, Database Center for

Life Sciences (Japan), and many others. For these organizations, there is an ongoing need to train new users and develop new workflows with CWL. Thus, improvements to CWL benefit multiple organizations and workflow systems that use the CWL standard.

One of the biggest barriers to CWL use and adoption is a perceived learning curve and challenges in usability, especially due to dense documentation and lack of examples.Our goal is to make it easier for users to create CWL workflows to support both the existing community and to expand adoption of CWL as a standard that enables the overarching mission of greater reproducibility in science.

Our objectives and proposed work:

1.Improve Formal and Informal Documentation

Organizations that have adopted CWL or are considering doing so have an ongoing need for education. This set of objectives focuses on expanding documentation including the User and Best Practice guides, training materials, templates, and the CWL website. The CWL website is the obvious first stop for people who want to learn more about CWL and therefore serves an important role as informal documentation. The expected outcome will be additional resources covering more of the CWLtopics and ecosystem as measured by new documentation pages, training materials, and code templates as well as a useful website as measured by analytics showing number of visits and timespent on the site.

1.1. Expand User Guide and create Best Practices guide.
1.2. Create CWL training materials.
1.3. Provide code templates demonstrating common patterns in CWL.
1.4. Update the CWL website.

2. Support Community and Ecosystem

Community support and feedback helps build strong communities which are essential to the success of open source projects. Our objectives will help encourage members to contribute their own expertise in the form of questions and answers, reusable community documents, software patches, and feedback. This will result in an expanded knowledge base, as measured by the number of questions and answers; expanding tool and workflow repositories, number of CWL documents and contributors;and the expanded role of community feedback as measured by percentage of requests for features and enhancements that are responded to and successfully resolved.

2.1. Support online discussion/Q&A forum.

2.2. Promote repositories of CommandLineTool and workflow documents.

2.3. Collect community feedback and incorporate into the roadmap.

3. Maintain Core CWL Specification and Tools

The core of CWL is the specification itself. It is the collaborative product of many people and organizations. The following objectives aim to add new features to CWL based on community feedback as measured by percentage of feature requests successfully resolved, and number of new conformance tests added to cover new features and/or resolve ambiguities in the specification.The reference implementation of the Common Workflow Language is “cwltool” written primarily by Peter Amstut (key personnel, 50% effort on this proposal.) Cwltool provides a feature complete implementation that other implementations can use as a basisf or comparison for correct behavior. Parts of cwltool are also used by several other CWL runners (Arvados, CWL-Airflow, and Toil), and it provides a number of other widely used features for working with CWL documents, such as validation. In addition to the reference implementation, there is an ecosystem of developer tools for working with CWL documents. This set of objectives involves the maintenance and development of these tools. The expected outcome of these objectives are successful resolution of bugs and pull requests, as measured by a reduction in the number of openbugs in Github repos for cwltool and other utilities mentioned.

3.1. Ongoing specification development.

3.2. Maintain the CWL test suite

3.3. Maintain the CWL reference runner cwltool.

3.4. Maintain tools for visualization and documentation of CWL.

Measuring Success

Our success will be measured in the completion of milestones and accessibility of CWL to the larger community. General progress on meeting objectives/milestones will be tracked via a weekly meeting. To track the usefulness of documentation and tooling, we will introduce feedback mechanisms (e.g. “is this page useful to you”), implement surveys, and usability testing. Our goal is to increase CWL usage and community involvement. To measure this, we will track visitors to the CWL website, GitHub metrics, workflows on Dockstore and Turwl.com, questions asked and answered in Discourse,downloads of cwltool, attendees of meetups and other CWL-based events, and various google search metrics.

7. Milestones and Deliverables (required):

List expected milestones and deliverables, and their expected timeline. Be specific and include (where possible) any goals for metrics the software project(s) are expected to reach upon completion of the grant (maximum of 500 words)

This timeline assumes a start date of July 1 2020 for the proposed activities.

Q1: July 2020 - September 2020
- Collect online community feedback (online FAQ, surveys, etc.) about documentation gaps and needs (Objective 2.3)
- Develop plan for updates for User Guide, training materials and code templates (Objective 1.1, 1.2 & 1.3)
- Implement web analytics for commonwl.org (Objective 1.4)
- Identify/Document different types of visitors to commonwl.org and their needs (Objective 1.4)
- Evaluate approaches to build a knowledge base of frequently asked questions (Objective 2.1)
- Identify organizations creating repositories of reusable CWL documents (Objective 2.2)

Q2: October - December 2020
- Rough draft due for feedback for updated User Guide, training materials, and code templates (Objective 1.1, 1.2 & 1.3)
- Release of code templates (Objective 1.3)
- Plan for website design update (Objective 1.4)
- Collect frequently asked questions into knowledge base (Objective 2.1)
- Rough draft guide on repositories for publishing and reusing CWL documents (Objective 2.2)

Q3: January - March 2021
- Update and release of User Guide major update (Objective 1.1)
- Develop plan for “Getting Started” guides for specific platforms (Objective 1.1)
- Create 1-page CWL Summary, collect User Stories and quotes for website (Objective 1.4)
- Attend Bioinformatics Open Source Conference (BOSC) present training materials and collect feedback (Objective 1.2 & 2.3)
- Develop a plan for improvement of CWL visualization and documentation tools (Objective 3.4)

Q4: April - June 2021

- Release “Getting Started” guides for specific platforms (Objective 1.1)
- Release of training materials (Objective 1.2)
- Release Website with new design (Objective 1.4)
- Release documentation on how to publish and reuse CWL documents (Objective 2.2)
- Develop tools for visualization and documentation of CWL documents (Objective 3.4)

Ongoing tasks throughout the period of work:
- Collect community feedback about CWL features (Objective 2.3)
- Specification development of new features (Objective 3.1)
- Maintenance of the CWL test suite (Objective 3.2)
- Maintenance of the CWL reference runner “cwltool” (Objective 3.3)

Deliverables:
- Updated User Guide
- Guides on testing and debugging workflows, FAIR best practices for CWL
- Free, open CWL training materials
- Free, open CWL code templates
- Guide on CWL integration with workflow and tool repositories
- Guides on running CWL in production on specific platforms
- Community based FAQ / knowledge base
- Summary document of community feedback, updated development roadmap
- Updated layout for CWL website (e.g. new color scheme, templates, navigational elements), analytics about visits to the page, and new informational content including downloadable 1-page summary, case studies, and user quotes
- New features in the CWL Specification, updated specification documents
- Reduced backlog of bug reports in CWL tools
- Added features for cwl-viewer and related CWL code documentation tools

8. Existing Support (required):

List current and recent financial support for the software project(s), including duration, amount in USD, and source of funding (maximum of 250 words)

Work on the CWL project has been supported by numerous organizations. See: https://www.commonwl.org

Work on the Arvados Free and Open Source Software (FOSS) project and CWL support in Arvados was funded primarily via VC, commercial and grant dollars at Curoverse, Veritas Genetics and, subsequently, Curii Corporation.

Current NIH grants to the PI:

$350,000 - NHGRI - HG009660 - INTEGRATED AND QUERYABLE INDIVIDUAL GENOME AND PHENOME DATA FROM THE PUBLIC DOMAIN - 1-SEP-2017 to 31-AUG-2020

$2,100,000 - NIGMS - GM119858 - ADAPTING THE BERKELEY BIG DATA ANALYTICS STACK TO GENOMICS AND HEALTH - 15 SEP-2017 31-AUG-2020

9. Landscape Analysis (required):

Describe the other software tools (either proprietary or open source) that the audience for this proposal is primarily using. How do the software projects in this proposal compare to these other tools in terms of size of user base, usage, and maturity? (maximum of 250 words)

A variety of proprietary services such as DNA Nexus and Seven Bridges Genomics as well as a dizzying set of "Do-It-Yourself" options exist to execute biomedical workflows at scale. As a result the vast majority of complex biomedical workflows cannot be ported from organization to organization and instead require significant massaging to run.

The CWL community began at the bioinformatics open source conference (BOSC) when engineers from Curoverse (subsequently Curii Corporation), Seven Bridges Genomics, the Galaxy Project and others joined together to make it easier for bioinformaticians to write workflows that could be consumed/edited/run by many vendors.

Since then the GA4GH, US FDA, US National Cancer Institute and others have included the CWL standard as a recommended way to write and run portable workflows.

Although CWL is widely adopted by end-users, it was intended to be a low level language that could be the basis of portability between many higher level / user level workflow descriptions. We believe that CWL has shown great promise in this regard.

With more investment in community building across numerous competing organizations--and especially by freeing up Peter Amstutz to do more work with the community rather than large Curii customers--we think we can improve the overall reproducibility of science generally. By building an open standard and open, welcoming community we believe that even bitter rivals can make workflow systems that use CWL to interoperate and are a joy to use for the bioinformatics community.

10. Diversity, Equity, and Inclusion Statement (required):

Advancing DEI is a core value for CZI and we are requesting information on your efforts in this area. Describe any efforts the software project(s) named in this proposal have undertaken to increase diversity, equity, and inclusion with respect to their contributors and audience. Please see examples from successful first cycle applications (maximum of 250 words)

We are building CWL for everyone. All participants in the community should be able to bring their full, authentic selves to the project. The CWL Code of Conduct promotes a harassment-free experience and does not tolerate harassment of participants in any form. CWL is a member project of Software Freedom Conservancy that promotes, improves, develops, and defends Free,Libre, and Open Source Software (FLOSS) projects. The project is formally managed by the elected CWL leadership team, however every-day project decisions are made by the CWL community which is open for participation by anyone.

Curii Corporation, employees Peter Amstutz (key-personnel, 50% time, CWL leadership team, wrote 80%+ of the CWL specification and "cwl-tool" reference implementation) and will be hiring and managing the community engineer to execute the proposed work. Curii is a women-led company with a strong commitment to diversity, parity, and inclusion. This includes a non-discrimination policy for race, ethnicity, religion, age, national origin, ancestry, sexual orientation, gender,gender identity, gender expression, military service, disability, or marital/parental status. Curii will focus on recruiting a diverse pool of applicants for the community engineer position. Recruitment will be a comprehensive process where we will search out and interview candidates from diverse groups and communities including coding bootcamps, women’s and minority tech groups, and through explicit advertisements on job sites and social media.

To further support people from groups underrepresented in software and technology, Curii will also be funding an Outreachy intern to support the CWL project and community alongside Peter and the community engineer.

Optional Attachments

Incomplete

Upload in PDF format; attachments should be uploaded in a combined single PDF. Include up to three pages of additional information. This section can include figures, charts and tables, references for the proposal, or any additional material in support of the proposal (maximum of three pages)

CV of Applicant

Completed - Feb 4 2020

Upload in PDF format; include current and recent employment, education history, and references to any major publications and software contributions (maximum of two pages)

2019

Filename: 2019.swz.CV.pdf **Size:** 93.5 kB

Budget Description

Completed - Feb 4 2020

Upload in PDF format; budgets can be uploaded in a combined single PDF or one PDF for each software project (one page per software project maximum)

- *Description of the costs to be funded by this grant at a high level and in narrative or tabular form, outlining costs for personnel (including names, if known), supplies, equipment, travel, meetings/hackathons/sprints, subcontracts, other costs, and up to 15% indirect costs (excluding equipment and subcontracts).*
- *Indirect costs are limited to up to 15% of direct costs and are included within the annual budget total. Indirect costs may not be assessed on capital equipment or subcontracts, but subcontractors may include up to 15% indirect costs of their direct costs.*
- *Budget should be requested in US dollars.*
- *International grantees must use all grant funds exclusively for activities conducted outside the United States of America. Travel expenses to the United States (including round-trip tickets) should not be covered from the requested grant funds.*
- *Application budgets must reflect the actual needs of the proposal. The Chan Zuckerberg Initiative will work closely with successful applicants to arrive at a mutually acceptable budget after review.*

CZI EOSS CWL Budget - DOMESTIC BUDGET

Filename: CZI_EOSS_CWL_Budget_-_DOMESTIC_BUDGET.pdf **Size:** 54.9 kB

Number of Open Source Software Project(s)

Completed - Feb 4 2020

Number of Open Source Software Projects

How many software projects are involved in this proposal that will be supported by this grant?

Each proposal can request funding for work that includes up to five open source software projects. If multiple software projects are involved, details must be entered for all of them.

If your proposal involves more than one software project, mark this step as complete so you can fill in details for additional projects. *You may need to scroll down using the scroll bar in the task menu on the left side of the page to see these tasks. To change the number of software projects indicated after marking as complete, click the three dots in the upper right of the task and select edit, which will re-open the task.*

1

Open Source Software Project #1: Details

Completed - Feb 4 2020

Software Project #1: Details

Provide details and metrics for each open source software project that will be supported by the grant to help us assess its impact and quality.

Software Project Details:

Complete the following table for Open Source Software Project #1 of your proposal. All URLs should be in the format https://example.com and only one primary link should be provided.

1. Software Project name (required)	Common Workflow Language (CWL)
2. Homepage URL (required)	https://www.commonwl.org
3. Hosting platform (required)	GitHub
4. Main code repository (e.g. GitHub URL) (required)	https://github.com/common-workflow-language/
5. DOI of major publication(s) describing software project (if applicable)	https://doi.org/10.6084/m9.figshare.3115156.v2

6. Social media handles (if applicable)	https://twitter.com/commonwl
7. Do you or software project key personnel have commit rights to the code repositories for this software project? (required)	Yes
8. Short description of software project (200 words maximum) (required)	The Common Workflow Language (CWL) project produces open standards for describing data analysis workflows and tools in a way that makes them portable and scalable across a variety of software and hardware environments including workstations, cluster, cloud, and high performance computing (HPC) environments. The CWL standards are designed to meet the needs of data-intensive science, such as Bioinformatics, Medical Imaging, Astronomy, Physics, and Chemistry and is well suited for describing large-scale workflows in cluster, cloud and high performance computing environments where tasks are scheduled in parallel across many nodes. CWL is developed by a multi-vendor working group consisting of organizations and individuals aiming to enable scientists to share data analysis workflows. The CWL project is maintained on GitHub and we follow the Open-Stand.org principles for collaborative open standards development. Legally CWL is a member project of Software Freedom Conservancy and is formally managed by the elected CWL leadership team. However, every-day project decisions are made by the CWL community which is open for participation by anyone.

List of Known Key Personnel:

Key personnel are people involved in the software project who will be supported by the grant if the application is successful.

Complete the following for the key personnel on the open source software project listed above (up to 5) (required); ***enter n/a if any field is not applicable****. Personnel to be hired that have not been identified at this time can be listed in the budget section. You may need to use the scroll bar at the bottom of the table to scroll right to view and to complete all fields. Alternatively, you can tab to move through and complete the fields.* ***To add another person/row (up to five), click the box at the end of the row.***

We collect (optional) gender statistics about the applicant team in order to study the gender diversity of teams applying for funding and to learn about diversity trends in the field. We will not consider these gender statistics as part of final grantmaking decisions.

	First name	Last name	Email address	Current employer	Job title	Developer username if applicable (e.g., GitHub handle)	Country of Residence	Add another person/row
1	Peter	Amstutz	peter.amstutz@curii.com	Curii Corporation	Principal Software Engineer	https://github.com/tetron	United States	✖
5	Sarah	Zaranek	swz@curii.com	Curii Corporation	CEO & Chief Scientist	https://github.com/swzcurov erse	United States	✖

Do any of the Key Personnel self-identify as one of the following? (optional)

Transgender, Woman / Female, Man / Male, Non-binary, Prefer to self-describe

Yes

How many of the listed key personnel self-identify as one of the below gender identities? (optional)

Please do not include requested information on a per person basis, we are looking for aggregated information.

Transgender	(No response)
Woman / Female	1
Man / Male	(No response)
Non-binary	(No response)
Prefer not to say	(No response)
Prefer to self-describe	(No response)

Software Project Metrics: Quality (required):

Complete for the open source software project listed above.

1. What is the software project license?

Permissive license (e.g. BSD 3-Clause, MIT, Apache 2.0)

2. What is the main programming language?

Python

3. Does the software project have a code of conduct?

Yes

3. Link (optional; format https://example.com):

https://github.com/common-workflow-language/common-workflow-language/blob/master/CODE_OF_CONDUCT.md

4. Does the software project have end-user documentation?

Yes

4. Link (optional; format https://example.com):

https://www.commonwl.org/user_guide/

5. Does the software project have an issue tracker?

Yes

5. Link (optional; format https://example.com):

https://github.com/common-workflow-language/common-workflow-language/issues

6. Does the software project have a community engagement / Q&A forum (self-hosted, on Stack Exchange etc.)?

Yes

6. Link (optional; format https://example.com):

https://cwl.discourse.group/

7. Does the software project have contribution / coding guidelines?

Yes

7. Link (optional; format https://example.com):

https://github.com/common-workflow-language/cwltool/blob/master/CONTRIBUTING.md

8. Are there examples or demo notebooks, scripts, and datasets?

Yes

8. Link (optional; format https://example.com):

https://www.commonwl.org/user_guide/

9. Is there a corresponding package available in a package manager (PyPi, CRAN, etc.)?

Yes

9. Link (optional; format https://example.com):

https://pypi.org/project/cwltool/

10. Does the software project support continuous integration for testing?

Yes

10. Comment (optional):

The CWL conformance tests are run against many of the CWL implementations. (1) https://ci.commonwl.org/
(2) https://github.com/common-workflow-language/common-workflow-language/blob/master/CONFORMANCE_TESTS.md

Software Project Metrics: Impact (optional):

Complete the following for the open source software project listed above. ***Providing metrics is optional and metrics can be approximate.*** *For each metric, please provide a source, clarify how the metric was computed, and/or provide any other comments. For monthly metrics, please provide data from the most recent month for which the corresponding metric is available.*

1. Complete the following table. List the number and explanation for each, if needed:

	Number	Comment
Scholarly paper(s) (including preprints) citing or mentioning the software project	185	https://www.zotero.org/groups/2294829/cwl/items
		(1) pypi: The CWL reference implementation (cwltool) has 1.4 million+ total downloads

Monthly users, if applicable (based on one or more of the following: monthly downloads from websites, monthly downloads from package managers, monthly unique requests for updates, etc.)	20000	(currently 55K per month) https://pepy.tech/project/cwltool Note: cwltool download required by the Arvados, CWL-Airflow, REANA, and Toil workflow runners/platforms but not by any Seven Bridges product, Cromwell, or IBM's CWLEXEC Besides these 7 implementations actively & publicly used in production there are at least 7 implementations in development. (https://www.commonwl.org/#Implementations) (2) CWL v1.0 GitHub repository :https://github.com/common-workflow-language/common-workflow-language/graphs/traffic - 960 "GitHub Stars" - 1,308 views, 457 unique visitors over last 2 weeks - 268 clones, 54 unique cloners over last 2 weeks (3) CWL Reference Implementation repository: - 159 "GitHub Stars" - 1,482 views, 351 unique visitors over last 2 weeks - 419 clones, 41 unique cloners over last 2 weeks (4) Used by 177 repositories, of which 27 are published packages: https://github.com/common-workflow-language/cwltool/network/dependents These are projects that depend on the CWL reference runner. Many more make use of the CWL standards themselves
		There are over 200 repositories on GitHub that depend on the CWL reference runner (cwltool).

Software projects that depend on the project (if applicable)	20	There are 20 plus other software projects that make use of the CWL standards themselves. https://github.com/common-workflow-language/cwltool/network/dependents
Monthly visitors to project's website, discussion forum (e.g. Stack Overflow), or similar	3700	Recently moved to https://cwl.discourse.group/ Biostars: About 168700 total views over 45.6 months. https://www.biostars.org/t/cwl/?sort=views&limit=all%20time&q=

2. Size of the largest potential user base:

	Number	Comment
Estimate the potential number of unique users who could adopt this project in the relevant field/discipline. Use as guidance the number of users of comparable projects, the number of papers published in the domain to which the project is applicable, number of labs able to adopt the project, etc.	over 100,000	Estimate the potential number of unique users who could adopt this project in the relevant field/discipline. Use as guidance the number of users of comparable projects, the number of papers published in the domain to which the project is applicable, number of labs able to adopt the project, etc.

3. List of upstream, downstream, or related software projects that the team is contributing to or receiving contributions from:

(1) Platforms supporting CWL implementations:

- Arvados: https://github.com/curoverse/arvados
- Toil: https://github.com/DataBiosphere/toil
- CWL-Airflow: https://github.com/Barski-lab/cwl-airflow

(2) Rabix Composer: https://github.com/rabix/composer

(3) Dockstore: https://dockstore.org/

(4) CWL language server Rabix Benten: https://github.com/rabix/benten

4. Additional metrics from project code repositories and package managers:

Provide a short description of any considerations or caveats we should be aware of when computing metrics (e.g. a recent change in the name or hosting of the repository), or any additional information you would like to share about the project's impact and quality. (maximum of 500 words)

Common Workflow Language is the "waist of the hourglass" as it is designed to support many user-facing tools above and execution environments below, to create network effects where each participant is able to benefit from the ecosystem.

Therefore, it is crucial to support the CWL standard and community to ensure a healthy ecosystem of tools, workflows, and interoperable production-scale supported implementations. Currently, there are 20,500 publicly available CWL tool and Workflow descriptions on GitHub (https://github.com/search?l=Common+Workflow+Language&q=cwlversion&type=Code) which speaks to the growing adoption of the CWL standard and the importance of the CWL mission to provide an open standard for describing analysis workflows and tools in a portable and scalable way.

If this proposal is funded, Peter Amstutz (CWL leadership team, wrote more than 80% of the CWL specification and "cwl tool") will have time to focus on the long tail of CWL users who are not paying Curii customers. Furthermore, the community engineer proposed is intentionally TBA with the plan to find a new developer with a strong background in open source and who will add more diversity to the project development team.

Other active contributors to the CWL project and members of the CWL Leadership Team, Michael R. Crusoe (CWL Project Lead) and others, are eager to see Peter have more time to devote to the community and have contributed to this proposal. Since their time is limited to work on the crucial areas of usability and community, they are excited that Peter and a new developer will tackle these issues.